

A New Perspective on Uncalibrated Photometric Stereo

Thoma Papadhimitri
University of Bern
Bern, Switzerland

thoma.papadhimitri@iam.unibe.ch

Paolo Favaro
University of Bern
Bern, Switzerland

paolo.favaro@iam.unibe.ch

Abstract

We investigate the problem of reconstructing normals, albedo and lights of Lambertian surfaces in uncalibrated photometric stereo under the perspective projection model. Our analysis is based on establishing the integrability constraint. In the orthographic projection case, it is well-known that when such constraint is imposed, a solution can be identified only up to 3 parameters, the so-called generalized bas-relief (GBR) ambiguity. We show that in the perspective projection case the solution is unique. We also propose a closed-form solution which is simple, efficient and robust. We test our algorithm on synthetic data and publicly available real data. Our quantitative tests show that our method outperforms all prior work of uncalibrated photometric stereo under orthographic projection.

1. Introduction

This paper is concerned with the task of recovering a surface from a collection of images captured from the same position but under different illuminations, a problem known in computer vision as *photometric stereo* [30]. This problem has been studied for over 3 decades and has brought some remarkable understanding of the relations between the photometry and geometry of objects. Arguably, one of the most remarkable results has been obtained for the more challenging problem of uncalibrated photometric stereo (UPS), where lights are unknown. It has been shown that one can solve UPS up to 3 parameters, the so-called Generalized Bas-Relief (GBR) ambiguity [5]. The main reason why that study had so much impact is because it did not exploit any additional assumptions about the scene, but, rather, it showed an intrinsic property of imaging surfaces. The key idea is the use of *integrability*, i.e., the constraint that relates a smooth function to its derivatives.

Since then most solutions to UPS have identified the 3 unknown parameters by employing different types of

heuristics, which require more or less stringent assumptions on the objects to be reconstructed or how images are captured [33, 25, 3, 11, 9, 26, 13, 35]. None of these methods, however, introduced a novel insight to the intrinsic properties of the imaging process as was done with the GBR ambiguity study.

In this paper, we introduce a novel insight in uncalibrated photometric stereo: **There is no ambiguity under the perspective projection model.** In other words, one can uniquely reconstruct the normals of the object and the lights given only the input images and the camera calibration (focal length and image center). Also, under the perspective projection, one can reconstruct the depth map from the normals up to a scale factor. The analysis is also paired with a robust and simple algorithm that achieves state-of-the-art performance in the reconstruction of the normal maps on real data.

2. Prior Work

Photometric stereo (orthographic/perspective). When the illumination directions and intensities are known, photometric stereo can be solved as a linear system. One of the simplifying assumptions that are typically used (together with the Lambertian image formation model, distant point light sources and ignoring shadows/inter-reflections) is the orthographic projection of the scene onto the image sensor. Although most prior work assumes an orthographic projection, the perspective projection has been shown to be a more realistic assumption: in [12], such model is considered in the case of known illumination sources distributed near the object, which makes the problem formulation different from the classical one which assumes distant point light sources. Another example is the work of Tankus and Kiryati [28], which is based on the perspective image formation model introduced in the shape from shading problem [29, 23, 7]. Here the partial derivatives of the natural logarithm of the depth map (instead of the depth map itself) are recovered via a closed-form solution based on image ratios. How-

ever, such formulation is sensitive to image noise, which may corrupt the estimated gradient map and eventually violate the integrability constraint. For this reason, in [17], the same perspective photometric stereo framework is used without the image ratios, but by imposing the integrability constraint in a numerical optimization scheme, similarly to [1]. Another related work is [31], which uses shape from shading with the same image formation model of [28] with an additional fall-off term. This work differs from classical photometric stereo approaches as the final reconstruction is the fusion of different shape from shading reconstructions, each obtained independently from a single image. Other approaches use the perspective projection in hybrid reconstruction algorithms to fuse photometric cues with active stereo [21, 34] or passive stereo [10, 19].

Uncalibrated photometric stereo (orthographic).

When no prior knowledge about the illumination, geometry, and reflectance is available, the problem is called uncalibrated photometric stereo. In such case, all prior work assumes an orthographic projection. Under this model the normals and lights can be obtained up to a 3×3 invertible matrix (9-parameters ambiguity) by computing a singular value decomposition of the data matrix (and by imposing that its rank be 3) [14]. If the integrability constraint is imposed, the ambiguity reduces to 3-parameters (the so-called GBR ambiguity) [5]. Several additional assumptions (on the geometry, the albedo and/or the lights) need to be imposed to the problem to fix the GBR ambiguity. For instance, in [3], this is done by minimizing the entropy of the albedo distribution. In [25], the GBR ambiguity is solved by grouping normals and albedo based respectively on image appearance and color. Other approaches consider specularities and glossy surfaces [9, 26], reflectance symmetries [27], inter-reflections [6], by considering the Torrance and Sparrow model [13] or additional constraints obtained by a ring-light configuration [35]. A recent work [11], solves the ambiguity by introducing the Lambertian Diffuse Reflectance (LDR) maxima and imposing normals and lights to be parallel at those locations. Other approaches instead exploit shadows [22, 16], dimensionality reduction [24], Helmholtz reciprocity principle [36], a general lighting environment [4] or look-up tables [15]. Only in [18], the perspective projection is considered to define notions of shadow equivalence and the Generalized Perspective Bas Relief Ambiguity (KGBR). Here, it is shown that we cannot obtain additional cues from shadows (under both orthographic and perspective projection) to solve the GBR ambiguity. However, the perspective projection was not incorporated into the image formation model and the corresponding analysis was not done.

Uncalibrated photometric stereo (perspective). In this work we focus for the first time on uncalibrated photometric stereo under perspective projection. Our main contribution is to show that under this model, the problem can be solved unambiguously by imposing only the integrability constraint and to devise a closed-form solution that is simple, efficient and robust.

3. Image Formation Model

A large set of problems in computer vision, such as photometric stereo, is concerned with measuring geometric and photometric properties of a scene from images. The first step in the design of solutions to such problems is to devise the image formation model, *i.e.*, how light reflected off the surfaces of objects is then measured on the camera sensor. A reasonable approximation of this process is the *perspective projection* model, where 3D points are mapped onto the 2D camera sensor by dividing the first two coordinates by the third one. However, because of the nonlinearity of this mapping, the perspective projection is often dropped in lieu of the more linear, and simpler to analyze, orthographic projection. As it turns out, the added nonlinearity of the perspective projection is also useful to disambiguate the uncalibrated photometric stereo solution (see sec. 5).

3.1. The Lambertian model

When objects in the scene are Lambertian and the illumination is a distant point light, the measured image irradiance due to a point in space $X = [x \ y \ z]^T \in \mathbb{R}^3$ can be written as

$$I = \rho \langle N, \frac{L}{\|L\|} \rangle e, \quad (1)$$

where ρ is the albedo of the surface at X , $N \in S^2$ is the normal at X , $\|\cdot\|$ denotes the length of a vector, $L/\|L\|$ is the unit-normal distant light direction and e is the light intensity. While the image irradiance I is measured at a pixel of the camera sensor, the normals and the albedo are related to objects in space. Thus, the next step is to define how a point X in space is mapped to a pixel on the image plane. One can define such mapping as a projection $\pi : \mathbb{R}^3 \mapsto \mathbb{R}^2$ where a pixel $(u, v) \doteq \pi(X)$. We further assume that the point X on the surface of the objects can be parametrized via the first two coordinates, *i.e.*, $X = [x \ y \ \hat{z}(x, y)]^T$ where $\hat{z}(x, y)$ is the so-called *depth map* of the scene. Then, we can also parametrize the normal map at X as the mapping

$$\hat{N}(x, y) \propto \begin{bmatrix} \nabla \hat{z}(x, y) \\ -1 \end{bmatrix} \quad (2)$$

where ∇ denotes the gradient with respect to x and y , and \propto means equality up to a scale factor.

In photometric stereo, K images are captured from the same viewing direction, but with different illumination directions $L_k \doteq [p_k \ q_k \ -1]^T \in \mathbb{R}^3$ and illumination intensities $e_k > 0$ for $k = 1, \dots, K$. Notice that in our notation we use the convention that both lights and normals point towards the camera (hence, the negative third coordinate).¹ Then, the k -th measured image irradiance becomes

$$I_k(u, v) = \hat{\rho}(x, y) \langle \hat{N}(x, y), L_k \rangle \frac{e_k}{\|L_k\|} \quad (3)$$

where

$$(u, v) = \pi \left(\begin{bmatrix} x \\ y \\ \hat{z}(x, y) \end{bmatrix} \right) \quad (4)$$

denotes the projection of X to the pixel coordinates (u, v) on the image plane. Because of the one-to-one correspondence between (u, v) and (x, y) , we can also introduce, similarly to work of Tankus and Kiryati [28], the following equivalent reparametrizations

$$\rho(u, v) \doteq \hat{\rho}(x, y), N(u, v) \doteq \hat{N}(x, y), z(u, v) \doteq \hat{z}(x, y)$$

which allow us to write

$$N(u, v) \propto \begin{bmatrix} \nabla z(u, v) \\ -1 \end{bmatrix} = \begin{bmatrix} z_x(u, v) \\ z_y(u, v) \\ -1 \end{bmatrix}. \quad (5)$$

Although cumbersome, this notation will make the analysis of different projections models much clearer.

3.2. Orthographic projection

Under the orthographic projection, the pixel coordinates satisfy $(u, v) = (x, y)$. Let

$$\begin{aligned} p &\doteq p(u, v) \doteq z_u(u, v) \\ q &\doteq q(u, v) \doteq z_v(u, v). \end{aligned} \quad (6)$$

Since $u \equiv x$ and $v \equiv y$, then taking derivatives in x is equivalent to taking derivatives in u and similarly for y and v ; hence,

$$\begin{aligned} z_u(u, v) &= z_x(u, v) \\ z_v(u, v) &= z_y(u, v). \end{aligned} \quad (7)$$

Then, the irradiance equation (3) becomes

$$I_k(u, v) = \rho(u, v) \frac{pp_k + qq_k + 1}{\|L_k\| \sqrt{p^2 + q^2 + 1}} e_k \quad (8)$$

where

$$N(u, v) \propto \begin{bmatrix} \nabla z(u, v) \\ -1 \end{bmatrix} = \begin{bmatrix} p(u, v) \\ q(u, v) \\ -1 \end{bmatrix}. \quad (9)$$

¹This notation is slightly different from what has been used in prior work, e.g., see [28].

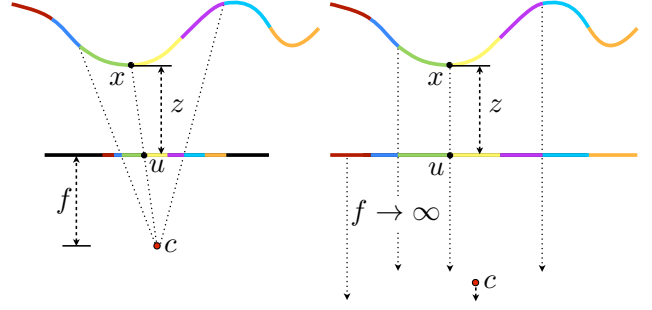


Figure 1. Left: The chosen perspective projection model. The reference frame is on the image plane where the coordinates (u, v) lie rather than on the camera center c . Right: When the focal length tends to ∞ then only the camera center shifts to $-\infty$ and the projection becomes orthographic.

3.3. Perspective projection

Under perspective projection we define the relationship between real world coordinates (x, y) and image coordinates (u, v) as

$$x(u, v) = u \frac{f + z(u, v)}{f}, \quad y(u, v) = v \frac{f + z(u, v)}{f} \quad (10)$$

or, vice versa, as

$$u(x, y) = x \frac{f}{f + \hat{z}(x, y)}, \quad v(x, y) = y \frac{f}{f + \hat{z}(x, y)} \quad (11)$$

where f is focal length of the camera. Notice that this model is slightly different from the one defined in [28]. The difference is illustrated in Fig. 1. As one can see, when changing the focal length we keep scene and image plane fixed and move only the camera center. This model was chosen so that when $f \rightarrow \infty$ the perspective projection converges directly to the orthographic one. By using the reparametrizations in eq. (5), the image irradiance equation (3) in the perspective case retains the same structure as with the orthographic projection. However, now $z_x(u, v) \neq z_u(u, v)$ and $z_y(u, v) \neq z_v(u, v)$. The form of $z_x(u, v)$ and $z_y(u, v)$ is obtained in the following proposition.

Proposition 3.1 *The derivatives of z with respect to x and y are*

$$\begin{aligned} z_x(u, v) &= \frac{f z_u}{f + z + u z_u + v z_v} \\ z_y(u, v) &= \frac{f z_v}{f + z + u z_u + v z_v} \end{aligned} \quad (12)$$

where z_u, z_v and z are all functions of (u, v) .

Proof. See appendix. ■

Let $\hat{p} \doteq \frac{z_u}{f+z}$ and $\hat{q} \doteq \frac{z_v}{f+z}$ and

$$p \doteq \frac{f \hat{p}}{1 + u \hat{p} + v \hat{q}}, \quad q \doteq \frac{f \hat{q}}{1 + u \hat{p} + v \hat{q}}. \quad (13)$$

Then, the normal map N can be written again as

$$N(u, v) \propto \begin{bmatrix} p(u, v) \\ q(u, v) \\ -1 \end{bmatrix} \quad (14)$$

yielding the same irradiance equation as in eq. (8).

Remark 1 Notice that the above analysis applied to the perspective projection equation of Tankus and Kiryati [28] yields the same formula that they obtain.

4. Uncalibrated Photometric Stereo

In the previous sections we have shown that both under the orthographic projection and the perspective projection one can obtain the same imaging equation simply by reparametrizing the normal map. Hence, one might wonder whether there are any fundamental differences between the two models. If we are given the light directions and intensities, the reconstruction of the normal map via the straightforward matrix pseudo-inverse $(\cdot)^\dagger$ is indeed identical in both cases

$$N(u, v) \propto \left[e_1 \frac{L_1}{\|L_1\|} \dots e_K \frac{L_K}{\|L_K\|} \right]^\dagger [I_1(u, v) \dots I_K(u, v)]. \quad (15)$$

However, when reconstructing the depth map from the normals the differences between the projection models become evident. In the case of the orthographic projection we can define $N(u, v) \doteq [N_1(u, v) \ N_2(u, v) \ N_3(u, v)]^T$ and then use

$$-\frac{N_1(u, v)}{N_3(u, v)} = z_u(u, v) \quad \text{and} \quad -\frac{N_2(u, v)}{N_3(u, v)} = z_v(u, v) \quad (16)$$

and then integrate the ratios to obtain $z(u, v)$, which coincides with $\hat{z}(x, y)$. In the case of the perspective projection we must instead use

$$\frac{N_1(u, v)}{N_3(u, v)} = \frac{-fz_u}{f+z+uz_u+vz_v} \quad \text{and} \quad \frac{N_2(u, v)}{N_3(u, v)} = \frac{-fz_v}{f+z+uz_u+vz_v}. \quad (17)$$

The integration to obtain $z(u, v)$ in this case leads to a different algorithm as previously observed [29, 28, 23] (also see sec. 6).

These, however, are not all the differences between the two models. To illustrate another fundamental difference, and the main contribution of this paper, we consider the uncalibrated photometric stereo problem, *i.e.*, when lights are not known. In general, all the ambiguities can be captured by a 3×3 (9 unknown parameters) invertible matrix Q [14] so that $\forall(u, v)$ and $k = 1, \dots, K$

$$\begin{aligned} I_k(u, v) &= B(u, v)^T S_k \\ &= \rho(u, v) N(u, v)^T Q^{-1} Q L_k e_k / \|L_k\| \end{aligned} \quad (18)$$

where $B(u, v) = Q^{-T} N(u, v) \rho(u, v)$ are the so-called pseudo-normals, and $S_k = Q L_k e_k / \|L_k\|$ are the pseudo-lights. Let us define the ambiguity matrix as $Q \doteq [Q_1^T \ Q_2^T \ Q_3^T]^T$, where Q_i , $i = 1, 2, 3$ are row vectors. It has been shown that the introduction of the surface integrability constraint, which exploits the relationship between the normal map and the derivatives of the depth map, in the orthographic projection case reduces the ambiguities to the so-called Generalized Bas-Relief (GBR) ambiguity [5], where $Q_1 = [1 \ 0 \ 0]$, $Q_2 = [0 \ 1 \ 0]$ and $Q_3 = [\mu \ \nu \ \lambda]$ for any μ, ν , and $\lambda \neq 0$. We show in the next section that the integrability constraint in the perspective projection case leads to the following outcomes:

1. There is no ambiguity
2. A closed-form unique solution in Q can be obtained.

Furthermore, as expected, when $f \rightarrow +\infty$ the solution converges to one of the GBR solutions.

5. Integrability Under Perspective Projection

The integrability constraint amounts to imposing that the order with which one takes the derivatives does not matter (as long as the depth map is smooth), *i.e.*, $z_{uv}(u, v) = z_{vu}(u, v)$. In the case of the orthographic projection this constraint is equivalent to $p_v(u, v) = q_u(u, v)$, so that the constraints directly applies to the normal map via $\frac{\partial}{\partial v} \frac{N_1(u, v)}{N_3(u, v)} = \frac{\partial}{\partial u} \frac{N_2(u, v)}{N_3(u, v)}$. In the case of the perspective projection, similarly to [17], we have the following result:

Proposition 5.1 *The integrability constraint $z_{uv} = z_{vu}$ holds if and only if the following constraint holds*

$$\hat{p}_v = \hat{q}_u. \quad (19)$$

Proof. Recall the definitions $\hat{p} \doteq \frac{z_u}{f+z}$ and $\hat{q} \doteq \frac{z_v}{f+z}$; then, we obtain

$$\hat{p}_v = \frac{z_{uv}(f+z) - z_u z_v}{(f+z)^2} \quad \hat{q}_u = \frac{z_{vu}(f+z) - z_v z_u}{(f+z)^2}, \quad (20)$$

and hence we have that $\hat{p}_v - \hat{q}_u = \frac{z_{uv} - z_{vu}}{f+z}$ which concludes the proof. ■

Since we are interested in using the integrability constraint to relate the entries of the normal map, we use eqs. (13) to write \hat{p} and \hat{q} as functions of p and q , *i.e.*,

$$\hat{p} = \frac{p}{f - up - vq} \quad \hat{q} = \frac{q}{f - up - vq} \quad (21)$$

so that

$$\begin{aligned} \hat{p}_v &= \frac{p_v(f - up - vq) - p(-up_v - q - vq_v)}{(f - up - vq)^2} \\ \hat{q}_u &= \frac{q_u(f - up - vq) - q(-p - up_u - vq_u)}{(f - up - vq)^2}. \end{aligned} \quad (22)$$

We are now ready to state the main result of this paper:

Theorem 5.2 *Given that the scene does not contain degenerate surfaces,² the integrability constraint in the case of perspective projection is sufficient to uniquely identify the normals N . Let B be the pseudo-normals defined as in eq. (18), then we have that $N \propto Q^T B$ with $Q = [\psi_1 \ \psi_2 \ \psi_3]^{-T}$, where the ψ_1, ψ_2, ψ_3 solve*

$$P \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} = 0 \quad (23)$$

and P is the $N \times 9$ perspective integrability matrix

$$P \doteq \begin{bmatrix} B_u^T \hat{B} & B_v^T \hat{B} & -\frac{u}{f} B_u^T \hat{B} - \frac{v}{f} B_v^T \hat{B} \end{bmatrix} \quad (24)$$

where B_u and B_v are the u and v derivatives of B respectively, and \hat{B} is the skew-symmetric matrix such that $\hat{B}A = B \times A$ for any vector A .

Proof. See Appendix. ■

The solution in Q can be readily found by computing the singular value decomposition of $P = U\Sigma V^T$, where U and V are orthogonal matrices, Σ is a diagonal matrix with positive entries and in decreasing order starting from the top-left corner, and by considering the last column vector in V , which corresponds to the smallest singular value in Σ .

Remark 2 *Notice that as $f \rightarrow +\infty$ the last 3 columns of the matrix P become 0 and hence we have 3 undetermined parameters as in the classical GBR solution. Indeed, the first six columns coincide with those obtained from the orthographic projection model as in the notation of [8].*

6. Depth Reconstruction Under Perspective Projection

Once the normal map N is reconstructed from the images, one can use eqs. (17) to recover the depth map z . Recall that such equations require

$$\frac{N_1(u,v)}{N_3(u,v)} = \frac{-fz_u}{f+z+uz_u+vz_v} \quad \text{and} \quad \frac{N_2(u,v)}{N_3(u,v)} = \frac{-fz_v}{f+z+uz_u+vz_v}. \quad (25)$$

Let $r_1(u,v) \doteq \frac{N_1(u,v)}{N_3(u,v)}$ and $r_2(u,v) \doteq \frac{N_2(u,v)}{N_3(u,v)}$. Then,

$$\begin{aligned} r_1z + (r_1u + f)z_u + r_1vz_v &= -fr_1 \\ r_2z + r_2uz_u + (r_2v + f)z_v &= -fr_2. \end{aligned} \quad (26)$$

In previous work [29, 28, 23] the above equations have been solved by using the substitution $g(u,v) \doteq \log(f + z(u,v))$ so that $g_u = \frac{z_u}{f+z}$ and $g_v = \frac{z_v}{f+z}$. Then, one obtains

$$g_u = \frac{-r_1}{f + ur_1 + vr_2} \quad g_v = \frac{-r_2}{f + ur_1 + vr_2} \quad (27)$$

²Degenerate surfaces are those that yield a rank of P in eq. (24) less than 8.

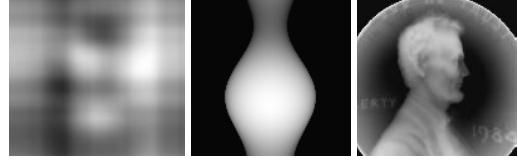


Figure 2. **Synthetic depth maps.** The 3 synthetic depth maps chosen for our synthetic experiments.

which can be readily solved via a Poisson solver [2] to yield g . Finally, the depth map is obtained by using $z(u,v) = e^{g(u,v)} - f$. Notice that because the function g can be reconstructed only up to a constant value c_0 , in our transformation this becomes a scale ambiguity, i.e., $z(u,v) = e^{g(u,v)+c_0} - f = s_0 e^{g(u,v)} - f$ where $s_0 = e^{c_0}$.

7. Experiments

7.1. Synthetic

To validate our analysis we run quantitative experiments (the Matlab code will be publicly available) on images synthetically generated under the perspective model with 3 given surfaces: **Random**, **Pot** and **Coin** (see Fig. 2). We consider 10 illumination conditions, a random albedo map, the center of the coordinate system at the center of the image and a focal length f equal to 100. We test the algorithm under several levels of additive Gaussian noise 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1% with a range of pixel values [0,1]. For each level we run our algorithm 20 times for different (random) configurations of lights and albedo. In Fig. 3 we compare the error plots (mean and standard deviation) of the mean angular error of the estimated normal maps obtained from calibrated photometric stereo and our uncalibrated solution. Notice that even in presence of no noise our implementation yields a non-negligible error. We find that this is due to numerical errors in the finite-difference approximations to the derivatives of the pseudo-normals. These errors become even more relevant when dealing with noise in the input images. Although we have used the same type of approximation employed by [3], we plan to address noise in photometric stereo in future work.

7.2. Real

We experimentally validate our method on publicly available real-world datasets: Cat, Buddha, Rock, Horse, Owl³ and Octopus, Redfish.⁴ We perform an image pre-processing step based on [20] as was done in [32, 11]. Our perspective model assumes the focal length and the camera center to be known. However, because no camera calibration is available for these datasets we fix the focal length

³<http://www.cs.washington.edu/education/courses/csep576/05wi/projects/project3/project3.htm>

⁴<http://neilalldrin.com/research>

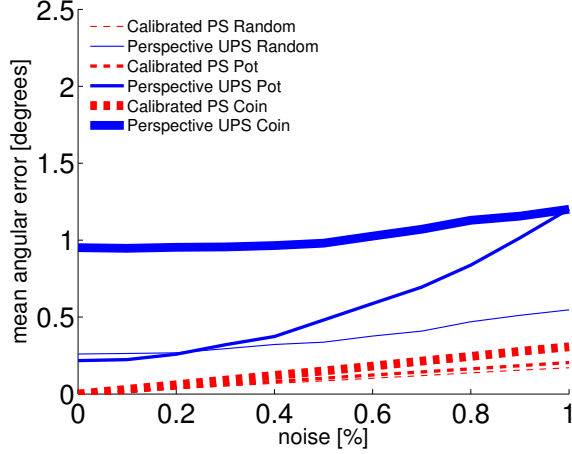


Figure 3. **Performance on synthetic data.** We show the plots for: Random, Pot and Coin mean angular errors. In the x -axis we show the percentage of noise added to the input images. In the y -axis we show the mean angular error of the reconstructed normal map compared to the ground truth.

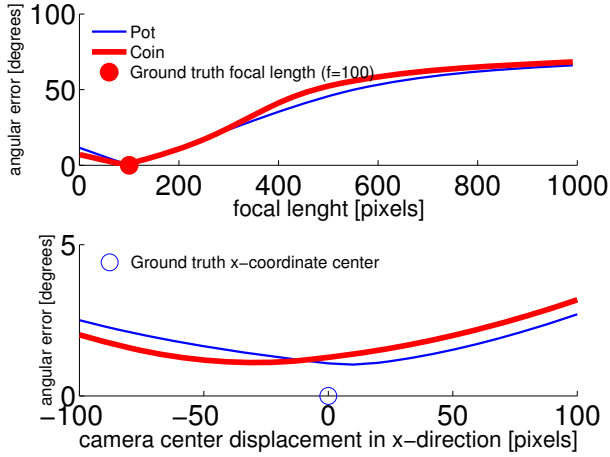


Figure 4. **Performance with different focal length and camera center.** We show the plots for: Pot and Coin mean angular errors. In the top x -axis we show the focal length (in pixels) used for the estimation of the normal map and in the bottom x -axis we show the displacement in x -direction (in pixels) of the camera center used for the estimation from the ground truth one. In the y -axis we show the mean angular error of the reconstructed normal map compared with the ground truth.

to 1. One can easily show that the reconstruction is obtained up to an unknown scaling, which we then fix based on the ground truth obtained from the photometric stereo results. We also manually search for the center of the camera (we choose the one that gives the best reconstruction). We compare our estimated normal maps (via the mean angular error with respect to the calibrated case) with the

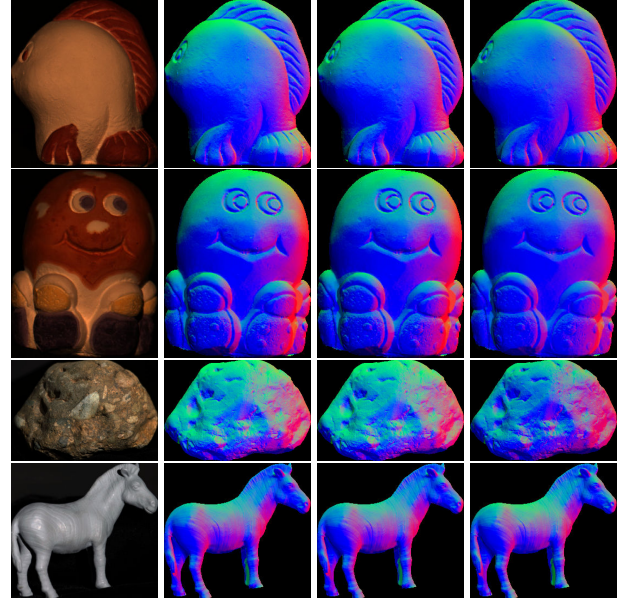


Figure 5. **Comparison of the normal maps (RGB).** First column: one of the input images. Second column: the normal map obtained from calibrated photometric stereo. Third column: the normal map obtained from our perspective uncalibrated method. Fourth column: the normal map obtained from the LDR orthographic uncalibrated method.

state-of-art in uncalibrated photometric stereo [3, 25, 11] under orthographic projection (see table 1). Such comparison shows the difference between adopting the orthographic or the perspective projection model (see sec. 3.3). Because our method does not depend on any assumptions about geometry and/or albedo and/or illumination, the errors are in the same range in all datasets. Also, notice that given the performance in the synthetic datasets (see sec. 7.1), most likely our errors are affected by the current numerical implementation. Our average running time (including the pre-processing step) for all the datasets is 7.40 seconds against 62 seconds [3], 10 minutes [25] and 13.5 seconds [11]. In Fig. 5 we show the depth maps obtained from the integration of the normals with a Poisson solver [2].

8. Conclusion

In this paper we provided analysis and experiments to demonstrate that uncalibrated photometric stereo under perspective projection can be solved unambiguously and robustly. Rather than exploiting heuristics about the scene, we used the lone integrability constraint and showed that it is sufficient to uniquely identify the light and normals given images satisfying the Lambertian model. We also demonstrated that our closed-form solution yields state-of-the-art performance on real data.

Table 1. Comparison with the Entropy Minimization (EM) method [3], the Self Calibrating Photometric Stereo (SCPS) method [25] and the Local Diffuse Reflectance (LDR) Maxima method [11]. We show the mean and the standard deviation of the angular error of the estimated normal maps.

Dataset (Nr. Images)	A/E	Redfish (5)	Octopus (5)	Rock (12)	Horse (12)	Buddha (12)	Cat (12)	Owl (12)
EM Method	μ	8.63	9.03	22.16	20.65	15.05	15.39	18.48
	σ	1.14	0.76	1.88	3.85	2.19	3.78	5.58
SCPS Method	μ	7.60	13.23	24.88	21.01	13.58	6.15	10.47
	σ	4.32	9.85	7.42	9.57	4.93	2.83	4.75
LDR Method	μ	5.60	6.64	11.61	4.80	4.98	5.37	6.63
	σ	0	0	0	0	0	0	0
Our method	μ	1.84	2.38	2.50	2.30	2.79	2.28	3.44
	σ	0	0	0	0	0	0	0

9. Appendix

9.1. Proof to Proposition 3.1

Proof. By using the perspective projection equations (11) we find the following system of equations

$$\begin{aligned}
 z_x(u, v) &\doteq \frac{\partial z(u(x, y), v(x, y))}{\partial x} \\
 &= z_u \frac{f}{f+z} - z_u z_x \frac{fx}{(f+z)^2} - z_v z_x \frac{fy}{(f+z)^2} \\
 z_y(u, v) &\doteq \frac{\partial z(u(x, y), v(x, y))}{\partial y} \\
 &= -z_u z_y \frac{fx}{(f+z)^2} + z_v \frac{f}{f+z} - z_v z_y \frac{fy}{(f+z)^2}.
 \end{aligned} \quad (28)$$

By grouping all terms in z_x on the left hand side of the first equation and similarly for the second equation, we obtain

$$\begin{aligned}
 z_x(u, v) \left(1 + z_u \frac{fx}{(f+z)^2} + z_v \frac{fy}{(f+z)^2} \right) &= z_u \frac{f}{f+z} \\
 z_y(u, v) \left(1 + z_u \frac{fx}{(f+z)^2} + z_v \frac{fy}{(f+z)^2} \right) &= z_v \frac{f}{f+z}
 \end{aligned} \quad (29)$$

or, equivalently,

$$\begin{aligned}
 z_x(u, v) \left(1 + z_u \frac{u}{f+z} + z_v \frac{v}{f+z} \right) &= z_u \frac{f}{f+z} \\
 z_y(u, v) \left(1 + z_u \frac{u}{f+z} + z_v \frac{v}{f+z} \right) &= z_v \frac{f}{f+z}
 \end{aligned} \quad (30)$$

from which we can immediately conclude the proof. ■

9.2. Proof to Theorem 5.2

Proof. Firstly, by using eqs. (22), the integrability constraint $\hat{p}_v = \hat{q}_u$ yields the following equation in p, q and their derivatives at each pixel (u, v)

$$f p_v - v q p_v + v p q_v - f q u + u p q_u - u q p_u = 0. \quad (31)$$

The above expression can be written more compactly as

$$\begin{bmatrix} p_v \\ q_v \\ 0 \end{bmatrix}^T \begin{bmatrix} 0 \\ -f \\ v \end{bmatrix} \times \begin{bmatrix} p \\ q \\ -1 \end{bmatrix} + \begin{bmatrix} p_u \\ q_u \\ 0 \end{bmatrix}^T \begin{bmatrix} -f \\ 0 \\ u \end{bmatrix} \times \begin{bmatrix} p \\ q \\ -1 \end{bmatrix} = 0 \quad (32)$$

where \times denotes the cross-product. Let us define $w_1 = [0 \ -f \ v]^T$ and $w_2 = [-f \ 0 \ u]^T$. Also, we substitute

$$\begin{bmatrix} p \\ q \\ -1 \end{bmatrix} = -\frac{Q^T B}{Q_3 B} \quad (33)$$

where we notice that Q_3 is the last row vector of Q , and the integrability constraint becomes

$$\begin{aligned}
 &\left(-\frac{Q^T B_v Q_3 B - Q^T B Q_3 B_v}{(Q_3 B)^2} \right)^T w_1 \times \left(-\frac{Q^T B}{Q_3 B} \right) \\
 &+ \left(-\frac{Q^T B_u Q_3 B - Q^T B Q_3 B_u}{(Q_3 B)^2} \right)^T w_2 \times \left(-\frac{Q^T B}{Q_3 B} \right) = 0.
 \end{aligned} \quad (34)$$

Since the terms $(Q_3 B)^2$ and $Q_3 B$ at the denominators are scalar numbers at each pixel (u, v) , we can multiply the whole equation for $(Q_3 B)^3$ and obtain

$$\begin{aligned}
 &(Q^T B_v Q_3 B - Q^T B Q_3 B_v)^T w_1 \times Q^T B \\
 &+ (Q^T B_u Q_3 B - Q^T B Q_3 B_u)^T w_2 \times Q^T B = 0.
 \end{aligned} \quad (35)$$

The cross-product between the vectors w_1 and $Q^T B$, and between w_2 and $Q^T B$, will be a vector orthogonal to $Q^T B$, which will therefore project to zero both $Q^T B Q_3 B_v \propto Q^T B$ and $Q^T B Q_3 B_u \propto Q^T B$; hence,

$$\begin{aligned}
 &(Q^T B_v Q_3 B)^T w_1 \times Q^T B \\
 &+ (Q^T B_u Q_3 B)^T w_2 \times Q^T B = 0.
 \end{aligned} \quad (36)$$

Again, we can divide both sides by the scalar $Q_3 B$ without affecting the equation, and arrive at

$$B_v^T Q w_1 \times Q^T B + B_u^T Q w_2 \times Q^T B = 0. \quad (37)$$

Now notice that the term $Q w_1 \times Q^T$ can be written as

$$\begin{bmatrix} 0 & -w_1^T Q_1^T \times Q_2^T & w_1^T Q_3^T \times Q_1^T \\ w_1^T Q_1^T \times Q_2^T & 0 & -w_1^T Q_2^T \times Q_3^T \\ -w_1^T Q_3^T \times Q_1^T & w_1^T Q_2^T \times Q_3^T & 0 \end{bmatrix}. \quad (38)$$

Define the column vectors $\phi_1 = Q_2^T \times Q_3^T$, $\phi_2 = Q_3^T \times Q_1^T$, $\phi_3 = Q_1^T \times Q_2^T$. Then, we have

$$Q w_1 \times Q^T = \begin{bmatrix} 0 & -w_1^T \phi_3 & w_1^T \phi_2 \\ w_1^T \phi_3 & 0 & -w_1^T \phi_1 \\ -w_1^T \phi_2 & w_1^T \phi_1 & 0 \end{bmatrix} = \kappa \widehat{Q^{-T} w_1} \quad (39)$$

where \hat{a} is the skew-symmetric matrix⁵ such that $\hat{a}b = a \times b$, and we notice that $Q^{-1} = \kappa^{-1}[\phi_1 \ \phi_2 \ \phi_3]$, with $\kappa = Q_1^T Q_2^T \times Q_3^T$; a similar derivation can be carried out for w_2 . Finally, we can rewrite the integrability constraint as

$$B_v^T (Q^{-T} w_1) \times B + B_u^T (Q^{-T} w_2) \times B = 0 \quad (40)$$

where κ has been removed from both terms. Since $a \times b = -b \times a$ we can write

$$B_v^T \hat{B} Q^{-T} w_1 + B_u^T \hat{B} Q^{-T} w_2 = 0. \quad (41)$$

Let the *perspective integrability matrix*

$$P \doteq \begin{bmatrix} B_u^T \hat{B} & B_v^T \hat{B} & -\frac{u}{f} B_u^T \hat{B} - \frac{v}{f} B_v^T \hat{B} \end{bmatrix} \quad (42)$$

⁵The skew symmetric matrix of a vector $a = [a_1 \ a_2 \ a_3]$ is defined as the matrix $\begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$.

be a tall $N \times 9$ matrix, where N is the number of pixels in the input images. Then, by writing $Q^{-T} = [\psi_1 \ \psi_2 \ \psi_3]$ eq. (41) can be written as

$$P \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} = 0. \quad (43)$$

If there are no degenerate surfaces in the scene then the rank of P will be 8 and hence the solution will be unique (up to a scaling factor). ■

References

- [1] A. Agrawal, R. Chellappa, and R. Raskar. An algebraic approach to surface reconstruction from gradient fields. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 - Volume 01*, ICCV '05, pages 174–181, Washington, DC, USA, 2005. IEEE Computer Society. 2
- [2] A. Agrawal and R. Raskar. What is the range of surface reconstructions from a gradient field. In *In ECCV*, pages 578–591. Springer, 2006. 5, 6
- [3] N. Alldrin, S. Mallick, and D. Kriegman. Resolving the generalized bas-relief ambiguity by entropy minimization. *Computer Vision and Pattern Recognition*, pages 1–7, 17–22 2007. 1, 2, 5, 6, 7
- [4] R. Basri and D. Jacobs. Photometric stereo with general, unknown lighting. *IEEE Conference on Computer Vision and Pattern Recognition*, pages II:374–381, 2001. 2
- [5] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *Int'l Journal of Computer Vision*, pages 35(1):33–44, 1999. 1, 2, 4
- [6] M. K. Chandraker, F. Kahl, and D. J. Kriegman. Reflections on the generalized bas-relief ambiguity. In *Conference on Computer Vision and Pattern Recognition*, pages 788–795, 2005. 2
- [7] F. Courteille, A. Crouzil, J. denis Durou, and P. Gurdjos. Towards shape from shading under realistic photographic conditions. In *Proceedings of the 17th International Conference on Pattern Recognition - ICPR 2004*, vol.2, pages 277–280, 2004. 1
- [8] O. Drbohlav. Towards uncalibrated photometric stereo for non-lambertian surfaces. *PhD Thesis: Center for Machine Perception (CMP), Czech Technical University, Prague, Czech Republic*, 2003. 5
- [9] O. Drbohlav and M. Chantler. Can two specular pixels calibrate photometric stereo? *Proc. of Int'l Conf. on Computer Vision*, pages 1850–1857, 2005. 1, 2
- [10] H. Du, D. Goldman, and S. Seitz. Binocular photometric stereo. In *Proceedings of British Machine Vision Conference (BMVC)*, 2011. 2
- [11] P. Favaro and T. Papadhimetri. A closed-form solution to uncalibrated photometric stereo via diffuse maxima. *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 1, 2, 6, 7
- [12] M. Galo and C. Tozzi. Surface reconstruction using multiple light sources and perspective projection. In *Image Processing, 1996. Proceedings., International Conference on*, volume 1, pages 309–312 vol.2, sep 1996. 1
- [13] A. Georgiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. *Proc. of Int'l Conf. on Computer Vision*, pages 816–823, 2003. 1, 2
- [14] H. Hayakawa. Photometric stereo under a light-source with arbitrary motion. *Journal of the Optical Society of America*, 11(11):3079–3089, 1994. 2, 4
- [15] A. Hertzmann and S. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, 2005. 2
- [16] H. P. K. Sunkavalli, T. Zickler. Visibility subspaces: uncalibrated photometric stereo with shadows. *European Conference on Computer Vision, Part II*, pages 251–264, 2010. 2
- [17] K. Kondo, Y. Mukaigawa, and Y. Yagi. Free-Form Mirror Design Inspired by Photometric Stereo. In *The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras - OMNIVIS*, 2008. 2, 4
- [18] D. J. Kriegman and P. N. Belhumeur. What shadows reveal about object structure. *Journal of the Optical Society of America*, 18:2001, 1998. 2
- [19] K. M. Lee and C.-C. Kuo. Shape from photometric ratio and stereo. *Journal of Visual Communication and Image Representation*, 7(2):155–162, 1996. 2
- [20] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *UIUC Technical Report UILU-ENG-09-2215*, 2009. 6
- [21] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH 2005)*, 24(3), Aug. 2005. 2
- [22] T. Okabe, I. Sato, and Y. Sato. Attached shadow coding: Estimating surface normals from shadows under unknown reflectance and lighting conditions. *Proc. of Int'l Conf. on Computer Vision*, pages 1693–1700, 2009. 2
- [23] E. Prados and O. Faugeras. Perspective shape from shading and viscosity solutions. In *International Conference on Computer Vision*, 2003. 1, 4, 5
- [24] I. Sato, T. Okabe, Q. Yu, and Y. Sato. Shape reconstruction based on similarity in radiance changes under varying illumination. *Proc. of Int'l Conf. on Computer Vision*, pages 1–8, 14–21 2007. 2
- [25] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1118–1125, 2010. 1, 2, 6, 7
- [26] P. Tan, S. Mallick, L. Quan, D. Kriegman, and T. Zickler. Isotropy, reciprocity and the generalized bas-relief ambiguity. In *Computer Vision and Pattern Recognition Conference*, pages 1–8, 17–22 2007. 1, 2
- [27] P. Tan and T. Zickler. A projective framework for radiometric image analysis. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2977–2984, 2009. 2
- [28] A. Tankus and N. Kiryati. Photometric stereo under perspective projection. *International Conference of Computer Vision*, pages I: 611–616, 2005. 1, 2, 3, 4, 5
- [29] A. Tankus, N. Sochen, and Y. Yeshurun. Shape-from-shading under perspective projection. *International Conference of Computer Vision*, 2004. 1, 4, 5
- [30] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 1
- [31] C. Wu, S. G. Narasimhan, and B. Jaramza. A multi-image shape-from-shading framework for near-lighting perspective endoscopes. *International Journal of Computer Vision*, 2009. 2
- [32] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Asian Conference on Computer Vision*, pages 703–717, 2011. 6
- [33] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. *Computer Vision and Pattern Recognition*, pages 158–164, 1997. 1
- [34] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2472–2479, june 2012. 2
- [35] Z. Zhou and P. Tan. Ring-light photometric stereo. In *Proceedings of the 11th European conference on Computer vision: Part II, ECCV'10*, pages 265–279, Berlin, Heidelberg, 2010. Springer-Verlag. 1, 2
- [36] T. Zickler, P. Belhumeur, and D. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *Int'l Journal of Computer Vision*, 49(2-3):215–227, 2002. 2